

The construction of customised nucleosomal arrays

Chenyi Wu, Christopher Read, John McGeehan and *Colyn Crane-Robinson

Biophysics Laboratories, School of Biology, University of Portsmouth, PO1 2DT, UK.

- Corresponding author:

Tel: +44-23-92842520, Fax: +44-23-92842053, colyn.crane-robinson@port.ac.uk

Highlights

- A method for construction of long, defined arrays of the 601 nucleosomal sequence
- High efficiency ligation of such long arrays
- Combination of the above can yield customised nucleosomal chromatin

Abstract

A simple, efficient and reliable method is demonstrated for cloning long tandem arrays of the 601 nucleosomal positioning sequence. Additionally, it is shown that such long arrays can be ligated together *in vitro* with high efficiency. By combining these two procedures it becomes straightforward to synthesise customised arrays that contain different (or variable) nucleosomal repeat lengths (NRLs) and monosome units bearing chemical modifications such as fluorophores, methyl groups or reaction sites. This is therefore an enabling technology for the *in vitro* study of chromatin structure and function.

Keywords: chromatin, nucleosome, 30 nm fibre, 601 sequence.

Introduction

The structure of the 30 nm chromatin fibre is a problem outstanding from the mid 70's. It was initially proposed (Finch and Klug, 1976) that the linker DNA is bent between adjacent nucleosomes in a simple one-start solenoid. Subsequently, several other models have been proposed, many of which have extended linkers that criss-cross the interior of the fibre (Staynov, 1983 & 2008; Williams *et al.*, 1986; Wong *et al.*, 2007). More recently, crystallographic and cryo-EM studies have shown short fibres in the form of a double helix, in support of what are known as two-start cross linker models (Schalch *et al.*, 2005; Scheffer *et al.*, 2011; Song *et al.*, 2014). However, structures of compact fibres reconstituted over a wide range of NRLs (nucleosome repeat lengths) reported by Robinson *et al.*, (2006) appear as interdigitated helices with undefined connectivity between nucleosomes. Resolving the path of the linker DNA in the fibre can be achieved by establishing the proximities of defined nucleosomes in long arrays, e.g. how close is a defined nucleosome (the i^{th}) to its near neighbours, the $(i+1)^{\text{th}}$, $(i+2)^{\text{th}}$ etc nucleosomes, using high resolution biochemical methods. For this, long nucleosomal arrays must be constructed with individually labelled monosomes at defined positions. Here we demonstrate a methodology capable of constructing such fully predefined chromatin fibres.

Tandem arrays of the 601 nucleosome positioning sequence (Lowary and Widom 1998) have been widely used for structural studies of the 30 nm chromatin fibre (Dorigo *et al.*, 2003; Dorigo *et al.*, 2004; Schalch *et al.*, 2005; Huynh *et al.*, 2005; Robinson *et al.*, 2006; Routh *et al.*, 2008; Kruithof *et al.*, 2009; Meng *et al.*, 2015; Song *et al.*, 2015). The strong histone octamer positioning power of the 601 sequence leads to arrays of nucleosomes with regular repeats, a feature presumed to facilitate folding into the 30 nm chromatin fibre (Huynh *et al.*, 2005) and maintain its regularity. Although natural chromatin may often lack such uniformity, investigation of chromatin higher order structure is facilitated by the availability of such regular *in vitro* assembled model systems.

Traditional methods of obtaining arrays of the 601 sequence often involve ligation of monomers, a process yielding multiple products, one of which is selected and then cloned. The length of the selected array is then determined relative to markers. Here we report a novel method of cloning 601 arrays in which the length of the array (the number of repeats) is predetermined and not selected from multiple products. We also show that such long cloned arrays can be ligated together with high efficiency. It follows that the combination of the two approaches can lead to the construction of customised arrays capable of chromatinisation with histones, i.e. defined chromatin templates can be assembled for *in vitro* structural and transcriptional assays. Such constructs are not limited to a single NRL and the spacing pattern of nucleosomes can be varied within the same array. Most importantly, natural nucleosome binding sequences and regulatory elements as well as chemical labels can be introduced at defined positions within the DNA template. Such a high degree of constructional flexibility should find wide application in chromatin research.

Results:

Cloning of 601 arrays by a doubling protocol

In a bacterial cloning vector (termed the 'Parental plasmid' in Figure 1) the 601 sequence is flanked with several restriction sites such that when linearized with DraIII and BstXI a free end forms on the insert,

generating the Acceptor plasmid. A separate Parental plasmid vector sample is then digested by BglI and BstXI, generating the Donor plasmid that has ends allowing its ligation into the Acceptor plasmid. In other words, the Parental plasmid construct is designed to generate both the Donor and Acceptor plasmids. After mixing the linearised Acceptor and the excised insert from the Donor, ligase is added to re-circularise and generate the Recombinant plasmid, which retains all the features of the Parental plasmid except that it now contains two inserts in tandem. Cells are then transformed with the Recombinant plasmid, which is harvested and becomes the Parental plasmid that allows the cycle to be repeated, with the result that the insert length is again doubled. In this way new constructs with increasing numbers of repeats can be obtained with high efficiency.

The working protocol is illustrated in Figure 1. Linearising the Parental plasmid yields the Acceptor plasmid with the insert having a CCC overhang available for ligation. In a separate aliquot the insert is fully excised to yield the Donor plasmid and is prevented from reforming the Parental plasmid by dephosphorylation. After clean-up to remove restriction enzymes and the small fragment released from the Acceptor plasmid, the products of the two reactions are mixed and ligated, the resulting Recombinant plasmid incorporating the free insert. Ligating the two non-palindromic complimentary overhangs generated by DraIII and BglI gives rise to a new recognition sequence for SmaI between the repeats, the original DraIII and BglI sites being lost.

In this example the new construct doubles the length of its insert and the method has so far been used successfully to multiply 601 monomer sequences up to 128-mer tandem repeats. It is important to appreciate that the Donor and Acceptor plasmids do not necessarily have to come from the same Parental plasmid and thus lead to doubling of the insert size. They could, for example, be of different nucleosome repeat lengths but they do need to be flanked by the appropriate restriction sites.

It is worth emphasising the relative simplicity of the protocol, in that no purification steps are required for the constituent inserts. Following enzymatic clean-up of Donor and Acceptor plasmids the total digests are simply mixed and ligase added. The Recombinant plasmid is immediately ready, either for excision for reconstitution into chromatin or for further use as a Parental plasmid.

Figure 2 illustrates this doubling protocol with the production of 16-mer and 32-mer repeats of a 167 bp 601 monomer sequence: partial SmaI digests allow one to count the number of repeats.

Making custom arrays in vitro

Once a collection of cloned repeats is obtained, individual inserts (containing different numbers of repeats or different NRLs) can be excised and used to generate synthetic constructs by stepwise *in vitro* ligation. Figure 3 shows two different inserts, a dimer and a monomer, both isolated by restriction with EcoNI and BstXI, that are then separately restricted with DraIII and with BglI to generate complimentary overhangs, neither insert being self-complimentary. After clean-up, the two inserts are mixed and dimerised into a single ligation product that inherits an intact BglI site on one end and an intact DraIII site on the other end and can therefore itself be restricted for further ligation by the same procedure. Figure 4 illustrates this process by ligating two 16-mer repeats of a 197 bp insert into a 32-mer: the high efficiency of the process shows that solution ligation of very large DNA duplexes is very effective.

It is important to appreciate that one ligation component, e.g. a mono-nucleosomal fragment, could be synthesised (or generated by PCR) to include embedded DNA modifications such as methylation,

fluorophores or reactive chemical groups. Provided the fragment carries the DraIII and BglI sites at its extremities it can be used as a building block in a longer array. In this way a labelled mononucleosome with Tag A can be introduced in the i^{th} position and, following ligation of a spacer, say a dinucleosome, a labelled monomer with Tag B introduced at the $(i+3)^{\text{th}}$ position. In this manner customised 601 arrays can be built to exact specifications.

Experimental methods

Figure 5 gives a detailed protocol for the key step of doubling the insert size shown diagrammatically in Figure 1. All plasmids are maintained in *Stbl4* cells (Invitrogen) that are designed for cloning unstable inserts such as tandem repeats. For maximal stability, cell cultures are grown and harvested at 30°C and transformation is performed by heat shock at 35°C. We have not experienced unwanted rearrangements in *Stbl4* cells over multiple generations, however rearrangements can occur in *DH5 α* cells as the host. Whichever cells are used, regular tests are necessary for quality assurance.

ACKNOWLEDGEMENTS The authors are particularly grateful to the Leverhulme Trust (RPG-2012-716) for the funding to carry out this project.

References

- Dorigo B, Schalch T, Kulangara A, Duda S, Schroeder RR, Richmond T (2004) Nucleosome arrays reveal the two-start organization of the chromatin fibre. *Science* **306**, 1571-1573.
- Dorigo, B, Schalch T, Bystricky K, Richmond TJ (2003) Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J. Mol. Biol.* **327**, 85-96.
- Finch JT, Klug A (1976) Solenoidal model for superstructure in chromatin, *Proc. Natl. Acad. Sci. U. S. A.* **73**, 1897–1901.
- Huynh VAT, Robinson PJJ, Rhodes D (2005) A method for the in vitro reconstitution of a defined 30 nm chromatin fibre containing stoichiometric amounts of the linker histone. *J. Mol. Biol.* **345**, 957-968.
- Kruithof M, Chien FT, Routh A, Logie C, Rhodes D, van Noort J (2009) Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fibre. *Nat. Struct. Mol. Biol.* **16**, 534-539.
- Lowary PT, Widom J (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19-42.
- Meng H, Andresen K, van Noort J (2015) Quantitative analysis of single-molecule force spectroscopy on folded chromatin fibres. *Nucleic Acids Res.* **43**, 3578-3590.
- Robinson PJJ, Fairall L, Huynh VAT, Rhodes D (2006) EM measurements define the dimensions of the “30-nm” chromatin fibre: Evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6506-6511.
- Routh A, Sandin S, Rhodes D (2008) Nucleosome repeat length and linker histone stoichiometry determine chromatin fibre structure. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8872-8877.

Schalch T, Duda S, Sargent DF, Richmond TJ (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**, 138-141.

Song F, Chen P, Sun D, Wang M, Dong L, Liang D, Xu RM, Zhu P, Li G (2014) Cryo-EM study of the chromatin fibre reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376-380.

Staynov DZ (1983) Possible nucleosome arrangements in the higher-order structure of chromatin. *Int. J. Biol. Macromol.* **5**, 3–9.

Staynov D, Proykova YG (2008) The sequentiality of nucleosomes in the 30 nm chromatin fibre. *FEBS J.* **275**, 3761-3771.

Williams SP, Athey BD, Muglia LJ, Schappe RS, Gough AH, Langmore JP (1986) Chromatin fibres are left-handed double helices with diameter and mass per unit length that depend on linker length. *Biophys. J.* **49**, 233-248.

Wong H., Victor JM, Mozziconacci J. (2007) An all-atom model of the chromatin fibre containing linker histones reveals a versatile structure tuned by the nucleosome repeat length. *PLoS ONE* **2**, e877.

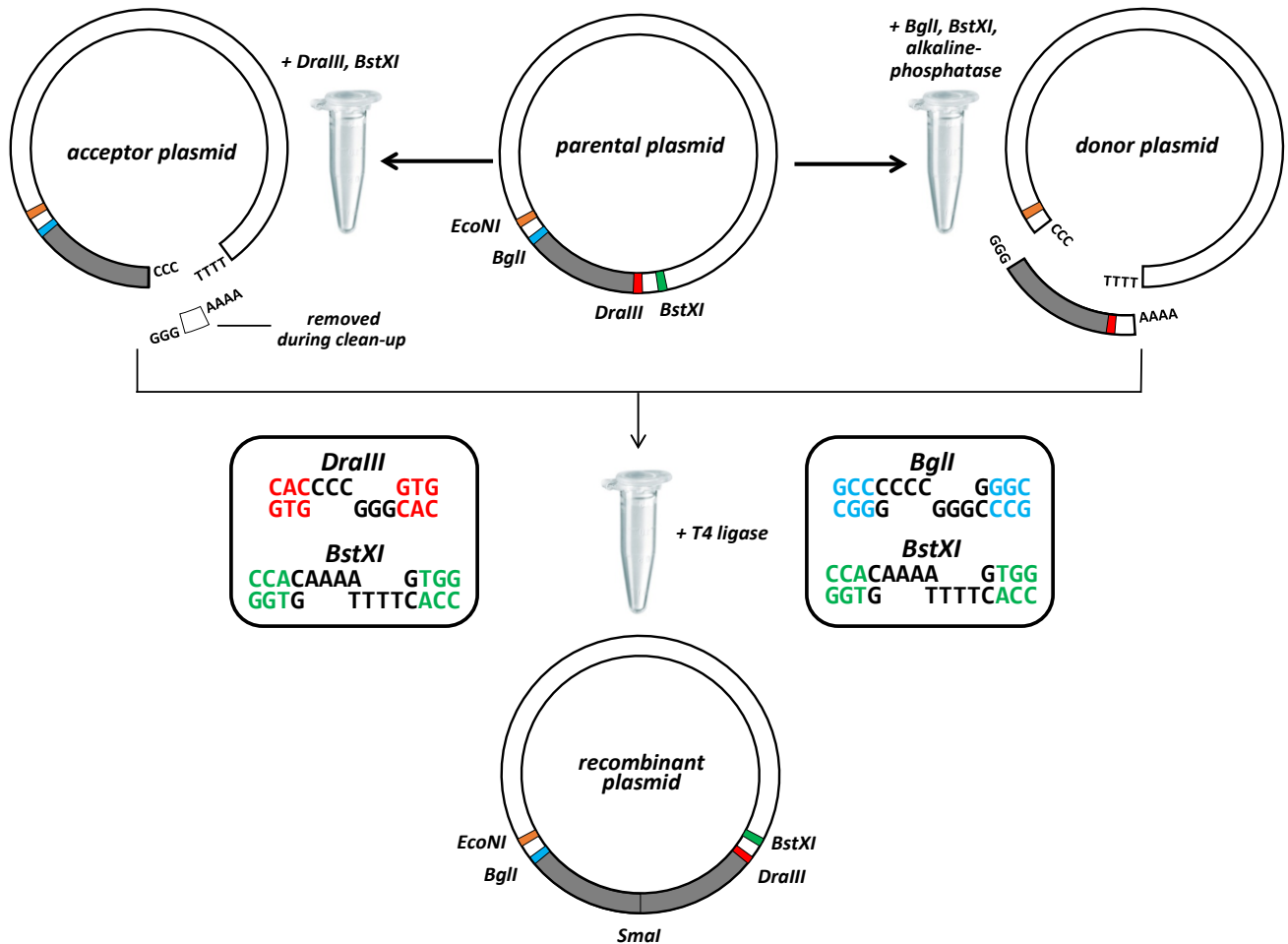


Figure 1 Expanding clones of 601 arrays by a doubling step. The Parental plasmid is subjected to two different and separate double restriction digests to generate an Acceptor plasmid capable of accepting an insert that is generated from the Donor plasmid. Phosphatase is also added to the Donor plasmid to prevent its subsequent re-ligation. Following clean-up (removal of small DNA fragments and enzymes) the two digests are mixed and ligase added. The doubled array has the original BstXI site restored and a SmaI site (CCCGGG) is generated between the two repeats. The Recombinant plasmid can then be used as the Parental plasmid for further length-doubling. The inset shows the restriction sites used. Black italic indicates recognition bases: coloured bases are selected to be suitable for subsequent cloning purposes.

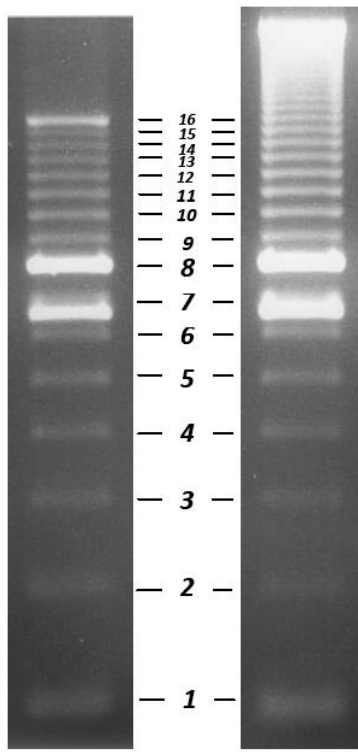


Figure 2 EFFECTIVE CLONING OF LONG REPEATS BY THE DOUBLING PROTOCOL OF FIGURE 1

Partial SmaI digests are used to check the lengths of tandem repeats, excised from Recombinant plasmids with EcoN1 and BstXI. Lane A is a digest of a 16-mer tandem repeat of a 167 bp 601 sequence obtained by doubling an 8-mer repeat using the protocol shown in Fig 1. Lane B is a digest of a 32-mer repeat likewise obtained by doubling the 16-mer. The two strong bands are vector products of 1373 and 1118 bp resulting from addition of BglI to prevent a strong uncut vector band (2.49 kb) overlapping the 16-mer insert (2.67 kb). The 1373 bp vector band corresponds to 8.2 repeats and the 1118 bp band to 6.7 repeats of the 167 bp monomer.

A: 16-mer B: 32-mer

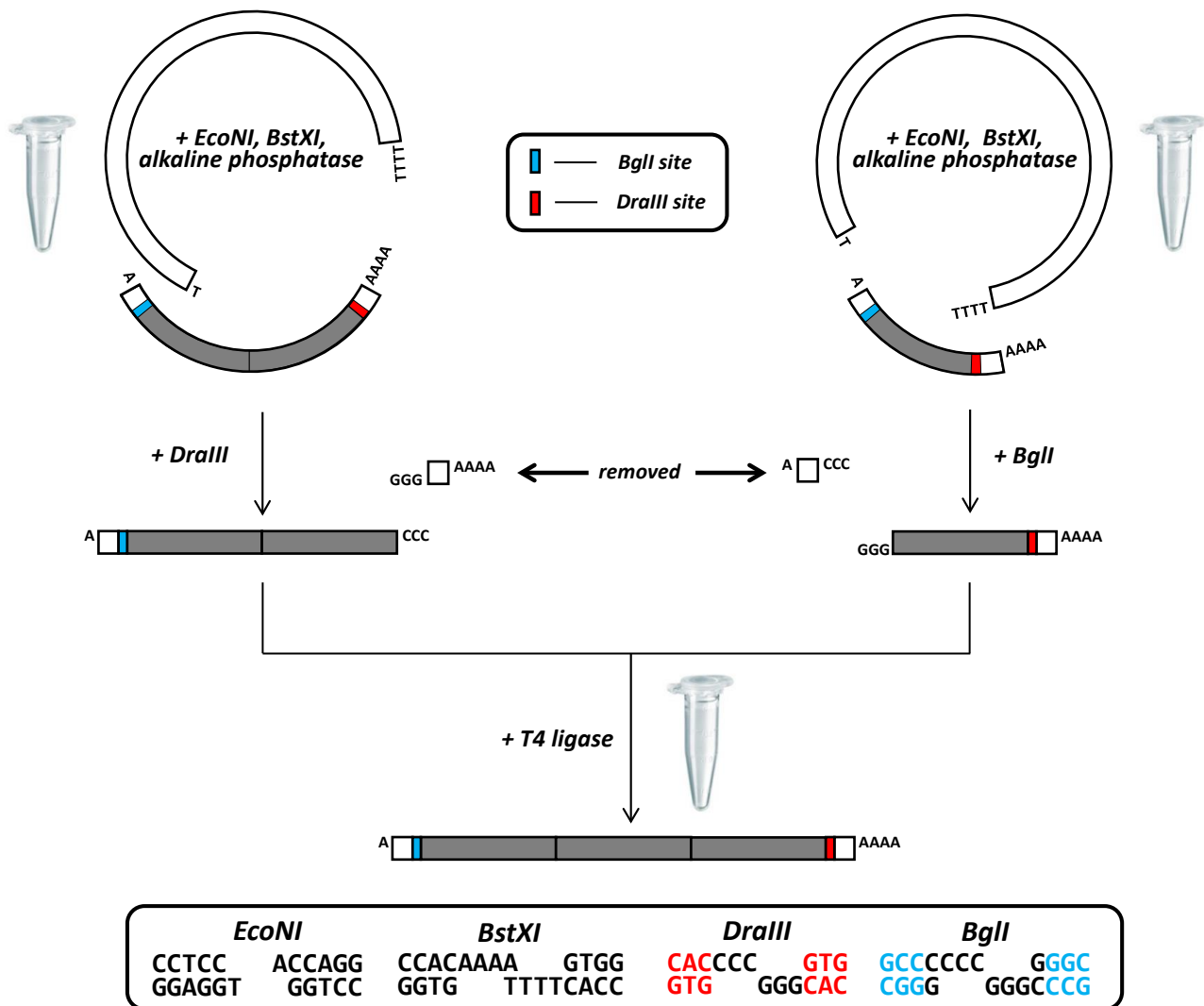


Figure 3 Making synthetic arrays by *in vitro* ligation. Individual inserts are excised from their respective plasmids by a double digest (*EcoNI* + *BstXI*) and phosphatase added to prevent re-circularisation. They are then separately restricted with *DraIII* and *BglI* to create complementary ends. After clean-up the contents of the two tubes are mixed and ligase added. High efficiency (i.e. good yield) results from the lack of self-complementarity in both inserts. The inset shows the restriction sites used. Black italic indicates recognition bases: coloured bases are selected for cloning purposes.

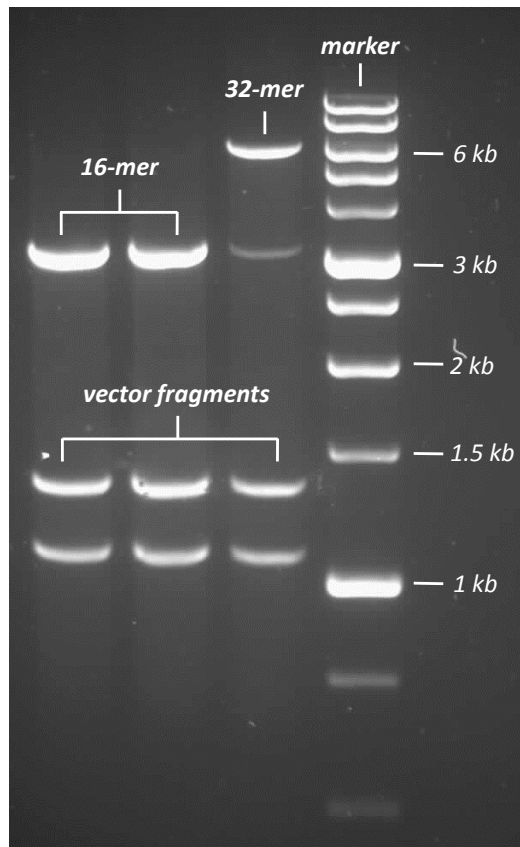
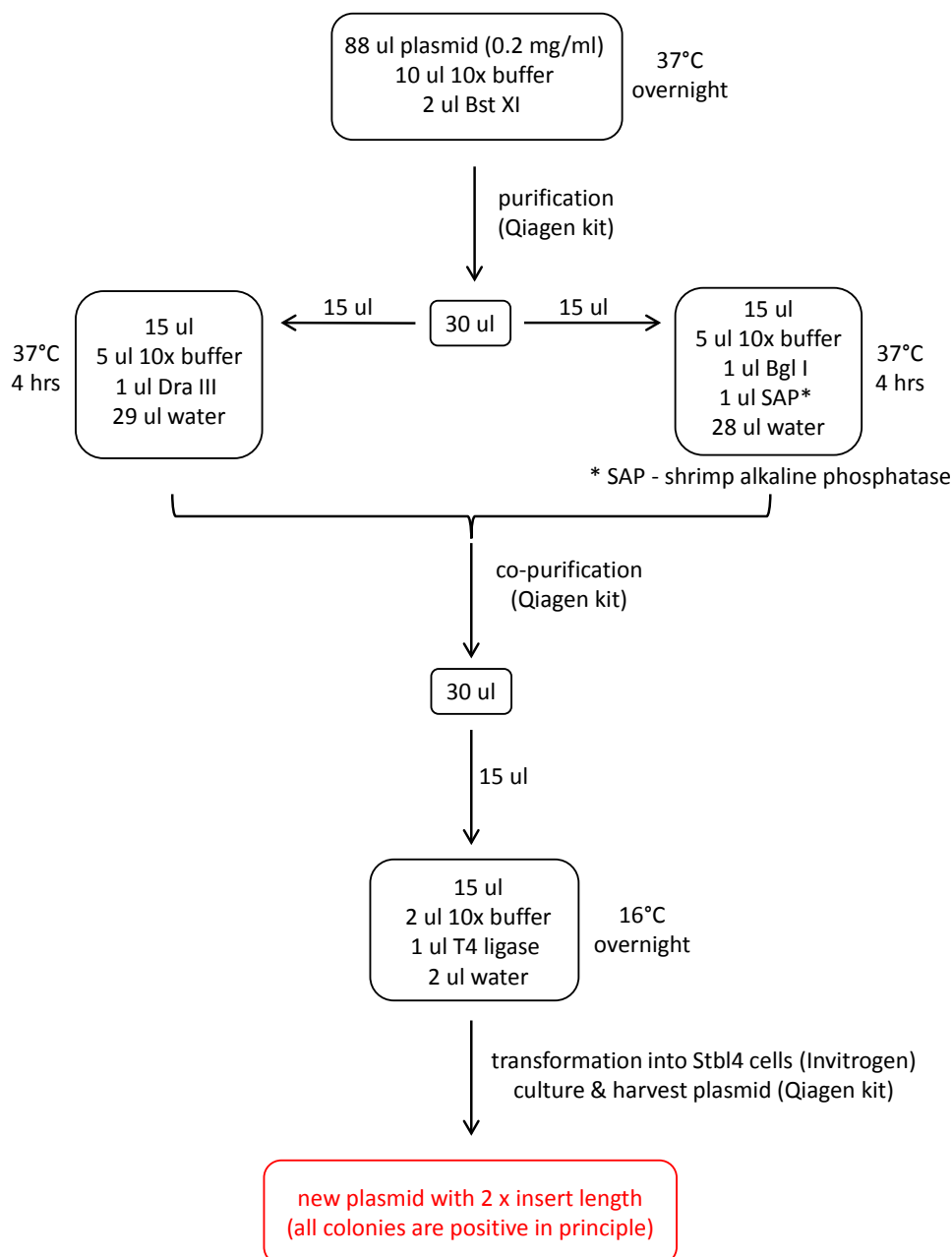


Figure 4 LIGATING LONG DNA FRAGMENTS IS A HIGHLY EFFICIENT PROCESS

An experiment ligating two 16-mers of a 197 bp repeat (3.15 kb) into a 32-mer (6.30 kb) using the protocol of Figure 3. Two aliquots of 16-mer inserts were excised from the plasmid with EcoNI and BstXI and phosphatase added to both. After clean-up these were separately restricted with DraIII and with BglI to generate complementary ends (see Figure 3): Lanes 1 and 2. The two solutions were mixed and ligase added. Lane 3 shows efficient formation of a 32-mer product. The weak 16-mer band remaining in Lane 3 is the result of imperfect 1:1 stoichiometry in the mix, not incomplete ligation. To prevent gel overlap of uncut vector and 16-mer inserts, the sample from the DraIII digest/ligation mix was also incubated with BglI before loading to digest the vector down to products of 1373 and 1118 bp (as in Figure 2).



1 x 197 NRL insert

————— **TCTAGAGATATCCCTCCACCAGGGGG****CCCCCGGGC**TGGACCCTATACGCGGCCGCCCTG
GAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACG
CGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATA
CATCCTGTGCATGTATTGAACAGCGAC**CACCCGTG**GG**CACAAAAGTGG**GATATC**GAATTC**—————

Figure 5 Detailed protocol for doubling the insert length (number of repeats) by the method shown in Figure 1. All enzymes and reaction buffers were purchased from New England Biolabs and used directly. The protocol can be accelerated by limiting the enzymatic reactions to 1 hr. For the primary vector construct, an insert corresponding to a single repeat was cloned between XbaI and EcoRI sites in pUC18. Lower panel: the 197 bp sequence is shown as an example. BstXI, BglI, and DraIII sites are coloured in green, blue, and red, respectively. The XbaI and EcoRI sites are at the two ends (in bold).